# Dynamic Carrier-Assisted Routing in Mobile Networks

Jie Wu

Department of Computer Science and Engineering

Florida Atlantic University

Boca Raton, FL 33431

*Abstract*— **We propose to use node mobility to enhance routing capability in a mobile network. A *dual-control planes* model is presented, which includes the traditional S(stationary)-plane for message routing among stationary nodes using connection-based message routing. In addition, we introduce a new control plane, M(mobile)-plane, for trajectory control of mobile nodes. A message routing consists of a sequence of control plane switches, alternating between store-and-forward in the S-plane and store-carry-forward in the M-plane. Intra-plane and inter-plane communication is performed through contacts. In the M-plane, unlike most existing works that focus on two extremes of spectrums (random movement and tightly controlled movement), our approach focuses on *loosely-controlled movement* of mobile nodes to assist message routing in a store-carry-forward paradigm. The challenge lies in adapting the degree of control to support a certain degree of flexibility, such as using dynamic node trajectory to maximize route sharing to minimize the number of carriers, while ensuring some desirable global properties, such as short average-moving-distance and a small number of relays to increase network capacity.**

## I. INTRODUCTION AND OBJECTIVE

In several emerging mobile networks including mobile ad hoc networks (MANETs), wireless sensor networks (WSNs), and the more recent delay-tolerant networks (DTNs) [3], one main issue centers around *whether mobility should be treated as a foe (undesirable) or a friend (desirable)* [13].

The traditional connection-based model used in the Internet (such as TCP/IP), MANETs, and WSNs, including most of the existing protocols (OSPF and RIP in the Internet and DSR, AODV and TORA in MANETs), is built on the premise that the underlying network is connected and views node mobility as undesirable. However, mobility is treated as a side issue through a simple recovery scheme. For example, a route disruption caused by node movement is made up by either route rediscovery or a local fix in a reactive approach. More recently, mobility has been identified as a serious threat to the traditional model [15]. The threat is mainly caused by asynchronous sampling of Hello messages and various protocol delays that result in an inconsistent global state. Several tolerant schemes have been proposed [14] as the first attempts to mask the effect of node movement and to construct a consistent global state.

Mobility as a means to improve system performance has been studied recently, including in-routing capability, location information dissemination, network capacity, and security. The more recent movement-assisted model tries to exploit node mobility for the routing process. The movement-assisted model (also called mobile infostation [4] or data-bus) typically follows a store-carry-forward paradigm, where a mobile node first stores the routing message, carries it while moving in a random or controlled fashion, and then forwards it to an intermediate node or the destination. This model is motivated by the following potential application areas: (1) In MANETs, the network capacity increases with resource-rich (in terms of processing, memory, and energy) mobile nodes, reducing the average number of relays in a routing process [5]. (2) In WSNs, the network lifetime is prolonged with mobile nodes in place of (bottleneck) stationary sensor nodes [12] to support relay and/or simple processing. Also, mobile sinks, such as Packbot [8] - an unmanned ground vehicle - can be used to collect data from disconnected and stationary sensors. (3) In DTNs, the network connectivity can be ensured using mobile nodes to connect various parts of a disconnected network [16].

However, most existing movement-assisted models in the DTN community focus on two extremes of the spectrum in mobility control: pure random movement (such as epidemic routing [11]) or total controlled movement (such as message ferrying [16]). The former may suffer unpredictable long delays with random movement in message delay while the latter may not offer sufficient flexibility in facing network dynamics with a fixed trajectory of mobile nodes. In addition, few existing works (except for [10]) deal with integrating movement-assisted models with connection-based models. Although several graph models have been proposed by incorporating time in edges of graphs [7] or in a new dimension to form a time-space view [9], these simplified models rely heavily on global information, especially the trajectory of mobile nodes.

We propose a *dual-control planes* model, which includes the traditional control plane, denoted as S(stationary)-plane, for message routing among fixed nodes using the conventional connection-based message routing as shown in Figure 1. In addition, we introduce a new control plane, denoted as M(mobile)-plane, for trajectory control of mobile nodes. In the general form, a message routing consists of a sequence of control plane switches, alternating between store-and-forward
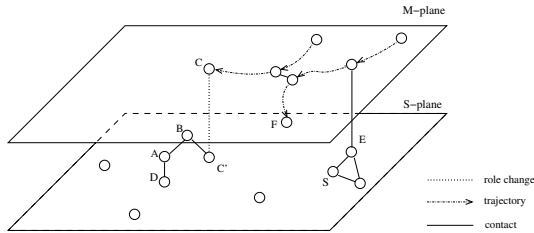
Fig. 1. Dual-control planes: M-plane for mobile control plane and S-plane for stationary control plane.



Fig. 2. Illustration of "eye" squares of (a) a $2 \times 2$ level-1 mesh, (b) a $4 \times 4$ level-2 mesh, (c) a $2^k \times 2^k$ level-$k$ mesh.

in the S-plane and store-carry-forward in the M-plane. Note that regular mobile nodes in the M-plane can still support connection-based message routing.

Intra-plane and inter-plane communication are performed through *contacts*. Here we assume that a contact of two nodes, stationary or mobile, is established once they are close to each other based on a given threshold. Note that we use the same notion of contact as in [2]. However, their usage and classification are different. A contact can be in the form of S-M (stationary-mobile, as between F and E in Figure 1), S-S (A and B in Figure 1), and M-M (C and F in Figure 1). The role of each node, stationary or mobile, can also be changed to further facilitate the routing process.

In the M-plane, our approach focuses on *loosely-controlled movement* of mobile nodes to assist message routing in a store-carry-forward paradigm. A subset of mobile nodes, called *carriers* similar to ferries in [16], is solicited to follow certain trajectories, not designed for a particular destination, to max-imize route sharing. The model supports not only relatively flexible node movement in the M-plane but also the role change of a node. That is, a mobile node, including a carrier, in the M-plane can be changed to a stationary node in the S-plane (such as C to $C'$ in Figure 1). In Figure 1, a possible route from $S$ to $D$ is $S \to E \to F \to C(C') \to B \to A \to D$. The challenge lies in adapting the degree of control to support a certain degree of flexibility, such as using dynamic node trajectory to maximize route-sharing to minimize the number of carriers, while ensuring some desirable global properties, such as a short average-moving-distance with respect to the Euclidian distance between the source and destination and a small number of relays to increase network capacity.

Here we provide some preliminary results of loosely con-trolled movement in the M-place. A simple proximity routing process in a 2-D Euclidian space with a stationary destination is proposed that achieves short average distance and controlled number of relays.

## II. PROPOSED MODEL

We consider the scenario in which the destination is sta-tionary and its location is known, and the routing space is restricted to the M-plane only which is the 2-D Euclidian space. Stationary destination is quite common, including the data sink in WSNs and a public access point leading to the Internet. Extensions will be discussed in our future work. In
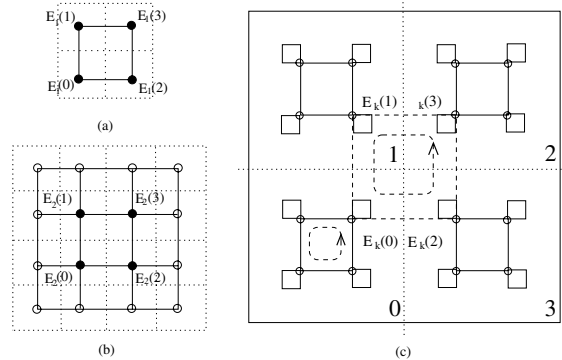
our basic model, we focus on designing loosely-controlled trajectory for carriers. Instead of applying pure random move-ment as in epidemic routing, or totally controlled movement as in message ferrying [16] with a single "track" for a ferry, our schemes use "multiple tracks" in the following way: multiple tracks are laid out in a global way and they interact through pre-determined locations (contacts), unlike mobile nodes in vehicle networks with random movement in a set of given tracks. Each carrier here follows the trajectory of a given track. However, each carrier does not have to follow one particular (global) track as in message ferry. Instead, it can move along multiple tracks (based on the need) through contacts. We assume the 2-D Euclidian space is partitioned into unit grids, each of which has a position $(x, y)$.

The global-layout multiple tracks use a hierarchical struc-ture in which tracks at the lowest level cover distinct regions. These tracks are connected to tracks at the higher levels through contacts. The location of contacts are defined based on the *eye theory* [1]. In this theory, special locations of relays (contacts) in a 2-D mesh, called eyes, are defined in such a way that a minimum total communication distance can be obtained if relays are done in these positions in a broadcast. All four grids in a $2 \times 2$ level-1 mesh are eyes of level-1, $E_1(j), j = 0, 1, 2, 3$. A level-$i$ mesh is partitioned into four level-$(i-1)$ submeshes, each of which has four eyes, $E_{i-1}(j)$. Eyes $E_i(j)$ are selected from sixteen $E_{i-1}(j)$s, specifically, eyes $E_i(j)$ are the four $E_{i-1}(j)$s that are the closest to the center of the level-$i$ mesh (see Figure 2). Each circular track is constructed by connecting four eyes in the same level. Each eye is a contact to a node in the S-plane, which stores messages between interface of multiple tracks. Figure 2 shows a level-$i$ track and a level-$(i-1)$ track and their contact at $E_k(0)$.

In dense mode with sufficient numbers of nodes in both the M-plane and the S-plane, mobile nodes have home tracks along which these nodes are circulated to fetch and drop messages at various contacts. Our routing process, called *logarithmic store-carry-forward routing*, is based on proximity routing through address matching. That is, a maximum of $\log n$ relays in an $n \times n$ square area is sufficient from one point to
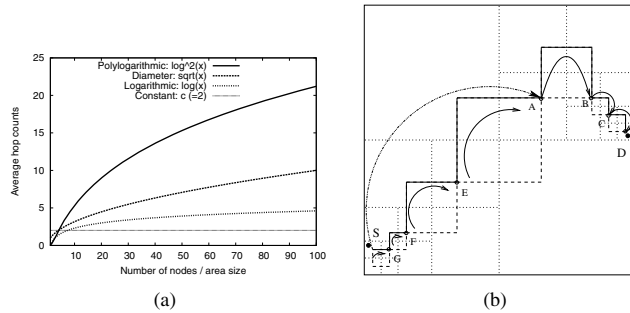
Fig. 3. (a) Paradigms with different hop counts and (b) a sample routing in eye space ($k = 4$).

another in the M-plane. While routing in the S-plane is faster, routing in the M-plane will increase the capacity of wireless networks [5]. This is mainly based on the result of Gupta and Kumar [6] which shows that when the number of nodes $m$ per unit area increases, the throughput per source-to-destination pair decreases as $\frac{1}{\sqrt{m}}$, due to the mutual interference of concurrent transmissions between nodes. Therefore, reducing the number of relays is equivalent to reducing the network density, and hence, increasing the network capacity. Note that in the traditional connection-based model as in the S-plane, the average hop count grows with the spatial diameter of the network (simply called *diameter-hop-count*) and while in the message ferry, two relays are needed (also called *constant-hop-count*) with significant delay. The proposed model, called *logarithmic-hop-count* shown in Figure 3 (a), is a compromise of the two in terms of capacity and delay in such an environment.

Our approach designs paths for both messages and carriers. This is because carriers of different levels follow their trajectories, while messages should be relayed several times by different carriers at contacts before reaching their destinations. The routing process has two unique features: (1) The "next hop" in the routing process corresponds to moving from the current contact to another through one or more transfers of different tracks. (2) The routing is conducted on a special space called the *eye space* built on top of the 2-D Euclidian space.

Because of the way eyes and tracks are defined, any two grid points can be connected through $k(= \log n)$ hops, i.e., $\log n$ numbers of track transfers via contacts. The eye space is defined in quaternary address: when the given square area is divided into four quadrants, each quadrant gets a quaternary address as shown in Figure 2 (c). The bottom left one is "0", and the others get "1", "2", "3" in clockwise order. This process repeats in each quadrant until unit grid is reached. Therefore, each grid point has an address $q = q_k q_{k-1} ... q_1$ in the 2-D eye space, with $q_k$ being the quaternary address in the highest level. The transfer from the 2-D Euclidian space to the 2-D eye space is straightforward: given a coordinate of a grid $(x, y)$ in the 2-D Euclidian space, with both $x$ and $y$ having $k$ bits. $q$ is its quaternary address, where $q_i$ is quaternary value for binary number $x_i y_i$. The quaternary address also reflects

the contact level for each grid. In general, a grid can be a contact at different levels. For example, in Figure 4, node E is contact for levels 1, 2, 3, and 4, node F is a contact for levels 1, 2, and 3, node G for levels 1 and 2, and node S for level 1.

Basically, each grid is a contact of level 1. If it is the contact of level $i - 1$ and $q_i + q_{i-1} = 3$, then it is also the contact at level $i$. For example, a grid point (5, 6) in a $2^4 \times 2^4$ 2-D Euclidian space has $x = 0101$ and $y = 0110$ when converting the number to binary. The quaternary address in eye space is then 0312. This grid is an eye (contact) at levels 1 and 2.

Suppose $q$ is the quaternary address of the destination, and $q'$ is the source. The first unmatched digit is $k$, therefore the "next hop" of the message is the eye of level $k$ with the address that begins with $q_k$. After the matching of first position, the routing space is reduced to one quarter of the original size. The same process is repeated until the last digit is matched. With the match of each digit, the message should go upstream/downstream level by level to reach the level of its "next hop". As in Figure 4 (a), the source $S$ is 0001 (meaning S in quarter 1 at level 1 and in quarter 0 at all other levels), and destination is 3222. When doing the matching, the first unmatched digit is 4, thus the next hop of this message should be the eye of level 4 with the address beginning with 3. Thus, according to the eye position rule, the next hop address is 3030, position $A$. To reach $A$ at level 4, the message may be relayed several times by carriers of different levels, from level 1 up to level 4. When the message reaches $A$ via track transfers at G, F, and E, it repeats the above matching process in a quarter region. The first unmatched digit of 3030 and 3222 is 3, thus, its next hop is the eye of level 3 with the address beginning with 32. It is node $B$, 3212. The next hop then is $C$, 3221, and the final hop is 3222.

In general, the match of the first digit usually corresponds to an "up phase" (moving up the level) and all the subsequent matches are "down phases" (moving down the level) (see Figure 3 (a)). In the up phase, each hop usually requires multiple track transfers while in the down phase, each hop requires at most one track transfer. So far, we implicitly assume carriers proactively circulate along their tracks. Carriers can be implemented on demand. As carriers are dynamically solicited from M-plane or S-plane they are dismissed when there is no message for the carrier. In this case, a carrier returns to its original grid and becomes a regular mobile node in the M-plane or a stationary node in the S-plane. For this purpose, each node keeps a variable $home$ that records its last position.

To reduce the potential problem of soliciting too many carriers, each carrier tries to carry multiple messages when possible and it stays at each contact before a given timer expires. Figure 4 (a) shows an example of routing in dense mode. Node $S_1$ is sending a message $m_1$ to $D_1$ and $S_2$ is sending $m_2$ to $D_2$. Message $m_1$ is carried to the top level, eye $B$, by carriers 1 and 2. Meanwhile, message $m_2$ arrives at another top level eye $G$, which solicits carrier 4 to carry this message. Carrier 4 arrives at $B$ before timer expiration and carries both $m_1$ and $m_2$. Message $m_1$ is dropped at the next eye $C$, where it is picked up by carrier 5 and delivered

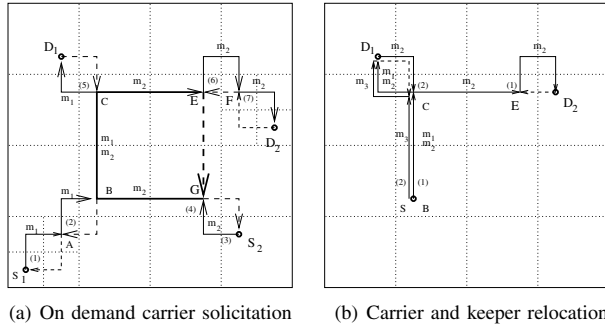(a) On demand carrier solicitation     (b) Carrier and keeper relocation

Fig. 4.  Store-carry-forward routing: (a) dense mode and (b) sparse mode.

to $D_1$. Message $m_2$ travels alone; it passes eyes $E$ and $F$ before it reaches $D_2$. A total of seven carriers are involved in the routing process. If there are no more messages to deliver, carriers will return to their home grid points and switch to regular mobile nodes in the M-plane or stationary nodes in the S-plane.

The situation is more involved in sparse mode with an insufficient number of nodes. In the case of an insufficient number of mobile nodes in the M-plane, a carrier can cover multiple tracks by combining them through their contacts. In the case of an insufficient number of stationary nodes in the S-plane, a carrier can change its status at the contact to become a stationary node to act as the storage place for other carriers to fetch and drop messages. As a solution, we propose a carrier and keeper reallocation scheme, which solicits carriers and keepers whenever there is a need. In addition, we try to recruit them from dense and low traffic regions and redistribute them to regions with high data traffic. By this we can also distribute nodes to balance load as a by-product.

Whenever a carrier reaches a contact without a keeper, it stays and acts as a new keeper that has received all the messages from the carrier. When a keeper has an expired timer but cannot find a mobile node in its grid, it solicits itself. That is, it acts as a new carrier to carry all messages in the previous keeper. We can imagine an extreme case with only two nodes in the network. The source node will travel all the way to deliver the data to the destination. Its status will alternate between keeper and carrier at each contact. Carriers are stateless and they "lose memory" after converting to keepers. When the converted keeper becomes a carrier again, its *home* grid is set to its current position instead of its original grid. This mechanism allows carriers to migrate to regions with needs and optimizes the node distribution over time.

Figure 4 (b) demonstrates the node redistribution process in sparse mode. Node $S$ is sending three messages: $m_1$ and $m_3$ to $D_1$, and $m_2$ to $D_2$. Nodes are densely deployed in eye $B$, but the other eye grids ($C$ and $E$) are empty. Assume message $m_2$ was generated before the timer of $m_1$ expires. The keeper in grid $B$ solicits carrier 1 to carry messages $m_1$ and $m_2$. As the next eye grid $C$ is empty, carrier 1 becomes its keeper and waits for a level 2 carrier (for $m_1$) and a level 3 carrier (for

$m_2$). After the timer expires, it solicits itself and delivers $m_1$ to $D_1$. Then it travels to $D_2$, alternating its role between keeper and carrier. Finally, it stops at eye $E$ and becomes a keeper. Message $m_3$ that is available at a later time is delivered by carrier 2, also solicited from grid $B$. After delivery, carrier 2 becomes a keeper of eye grid $C$. After the node redistribution, faster delivery is expected to destinations $D_1$ and $D_2$.

Our general plan of future work will include several other critical issues: (1) How to support other methods of track allocation, such as random allocation, without losing the desirable global property - a relatively small number of relays in the M-plane. (2) How to use the basic one-to-one unicasting to support collective communication, including multicast, broadcasting, anycast, gather, and gossip. (3) How to integrate the routing method in the M-plane with the traditional method in the S-plane. (4) How to deal with routing with mobile destination. That is, the location of destination cannot be precisely known at the routing source. (5) How to extend the dual control space model to other models, including high-dimensional and non-Euclidian space. (6) How to conduct an empirical study on the performance of proposed models and their effect on various global goals.

## REFERENCES

[1]  S. Cang and J. Wu.  Minimizing total communication distance of a broadcast on mesh and torus networks.  In *Proc. of First Merged Symposium IPPS/SPDP*, 1998.

[2]  V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss. Delay-tolerant network architecture. *Internet draft: draft-irtf-dtnrg-arch-04.txt, DTN Research Group*, 2006.

[3]  K. Fall.  A delay-tolerant network architecture for challenged internets. In *Proc. of ACM SIGCOMM*, 2003.

[4]  D. J. Goodman, J. Borras, N. B. Mandayam, and R. D. Yates. INFO-STATIONS: A new system for data and messaging services. In *Proc. of IEEE VTC*, 1997.

[5]  M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *Proc. of IEEE INFOCOM*, 2001.

[6]  P. Gupta and P. R. Kumar.  The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, 2000.

[7]  S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proc. of SIGCOMM*, 2004.

[8]  A. Kansal, A. A. Somasundara, D. D. Jea, M. B. Strivastava, and D. Estrin.  Intelligent fluid infrastructure for embedded networks.  In *Proc. of ACM MobiSys*, 2004.

[9]  S. Merugu, M. Ammar, and E. Zegura.  Routing in space and time in networks with predictable mobility. *Technical report: GIT-CC-04-07, College of Computing, Georgia Tech.*, 2004.

[10]  D. Nain, N. Petigara, and H. Balakrishnan. Integrated routing and storage for messaging applications in mobile ad hoc networks. In *Proc. of WiOpt*, 2003.

[11]  A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. *Technical report, Duke University*, 2002.

[12]  W. Wang, V. Srinivasan, and K. C. Chua. Using mobile relays to prolong the lifetime of wireless sensor networks.  In *Proc. of ACM MobiCom*, 2005.

[13]  J. Wu. Mobility in ad hoc wireless networks: A villain or a friend? In *Proc. of Third International Workshop on Mobile Distributed Computing, in conjunction with IEEE ICDCS*, 2005.

[14]  J. Wu and F. Dai. Mobility management and its applications in efficient broadcasting in mobile ad hoc networks. In *Proc. of IEEE INFOCOM*, 2004.

[15]  J. Wu and F. Dai. Mobility-sensitive topology control in mobile ad hoc networks. In *Proc. of IEEE IPDPS*, 2004.

[16]  W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transport ferries in a delay-tolerant network.  In *Proc. of IEEE INFOCOM*, 2005.